



"Out of the Can"

A Draft Genome Assembly, Liver Transcriptome, and Nutrigenomics of the European Sardine, *Sardina pilchardus*

Machado, André M.; Tørresen, Ole K.; Kabeya, Naoki; Couto, Alvarina; Petersen, Bent; Felicio, Mónica; Campos, Paula F.; Fonseca, Elza; Bandarra, Narcisa; Lopes-Marques, Mónica; Ferraz, Renato; Ruivo, Raquel; Fonseca, Miguel M.; Jentoft, Sissel; Monroig, Óscar; da Fonseca, Rute R.; Castro, L. Filipe C.

Published in:
Genes

DOI:
[10.3390/genes9100485](https://doi.org/10.3390/genes9100485)

Publication date:
2018







Document version
Publisher's PDF, also known as Version of record

Document license:
[CC BY](#)

Citation for published version (APA):
Machado, A. M., Tørresen, O. K., Kabeya, N., Couto, A., Petersen, B., Felicio, M., Campos, P. F., Fonseca, E., Bandarra, N., Lopes-Marques, M., Ferraz, R., Ruivo, R., Fonseca, M. M., Jentoft, S., Monroig, Ó., da Fonseca, R. R., & Castro, L. F. C. (2018). "Out of the Can": A Draft Genome Assembly, Liver Transcriptome, and Nutrigenomics of the European Sardine, *Sardina pilchardus*. *Genes*, 9(10), 1-13. [485].
<https://doi.org/10.3390/genes9100485>

Article

"Out of the Can": A Draft Genome Assembly, Liver Transcriptome, and Nutrigenomics of the European Sardine, *Sardina pilchardus*

André M. Machado ^{1,†} , Ole K. Torresen ^{2,†} , Naoki Kabeya ^{3,†} , Alvarina Couto ^{1,4}, Bent Petersen ^{5,6} , Mónica Felício ⁷, Paula F. Campos ^{1,4}, Elza Fonseca ^{1,8}, Narcisa Bandarra ⁷, Mónica Lopes-Marques ¹, Renato Ferraz ^{1,9}, Raquel Ruivo ¹, Miguel M. Fonseca ¹, Sissel Jentoft ^{2,10,*} , Óscar Monroig ^{11,*}, Rute R. da Fonseca ^{4,12,*}  and L. Filipe C. Castro ^{1,8,*}

¹ Interdisciplinary Centre of Marine and Environmental Research (CIIMAR), University of Porto, 4450-208 Matosinhos, Portugal; andre.machado@ciimar.up.pt (A.M.M.); alvarinacouto@gmail.com (A.C.); campos.f.paula@gmail.com (P.F.C.); fonseca.ess@gmail.com (E.F.); monicaslm@hotmail.com (M.L.-M.); renatobarbosferraz@gmail.com (R.F.); ruivo.raquel@gmail.com (R.R.); mig.m.fonseca@gmail.com (M.M.F.)

² Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, 0371 Oslo, Norway; o.k.torresen@ibv.uio.no

³ Department of Aquatic Bioscience, The University of Tokyo, Tokyo 113-8654, Japan; naokikby@g.ecc.u-tokyo.ac.jp

⁴ The Bioinformatics Centre, Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark

⁵ Natural History Museum of Denmark, University of Copenhagen, DK-1350 Copenhagen, Denmark; bent.petersen@snm.ku.dk

⁶ Centre of Excellence for Omics-Driven Computational Biodiscovery, Faculty of Applied Sciences, Asian Institute of Medicine, Science and Technology, Kedah 08000, Malaysia

⁷ Portuguese Institute for the Sea and Atmosphere, (IPMA), 1749-077 Lisbon, Portugal; monicafelicio124@hotmail.com (M.F.); narcisa@ipma.pt (N.B.)

⁸ Department of Biology, Faculty of Sciences, University of Porto, 4099-022 Porto, Portugal

⁹ Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, 4099-022 Porto, Portugal

¹⁰ Centre for Coastal Research, Department of Natural Sciences, University of Agder, 4630 Kristiansand, Norway

¹¹ Instituto de Acuicultura Torre de la Sal, Consejo Superior de Investigaciones Científicas (IATS-CSIC), 12595 Ribera de Cabanes, Spain

¹² Center for Macroecology, Evolution, and Climate, Natural History Museum of Denmark, University of Copenhagen, DK-2100 Copenhagen, Denmark

* Correspondence: sissel.jentoft@ibv.uio.no (S.J.); oscar.monroig@csic.es (Ó.M.); rfonseca@snm.ku.dk (R.R.d.F.); filipe.castro@ciimar.up.pt (L.F.C.C.)

† These co-first authors contributed equally to this work.

Received: 4 September 2018; Accepted: 2 October 2018; Published: 9 October 2018



Abstract: Clupeiformes, such as sardines and herrings, represent an important share of worldwide fisheries. Among those, the European sardine (*Sardina pilchardus*, Walbaum 1792) exhibits significant commercial relevance. While the last decade showed a steady and sharp decline in capture levels, recent advances in culture husbandry represent promising research avenues. Yet, the complete absence of genomic resources from sardine imposes a severe bottleneck to understand its physiological and ecological requirements. We generated 69 Gbp of paired-end reads using Illumina HiSeq X Ten and assembled a draft genome assembly with an N50 scaffold length of 25,579 bp and BUSCO completeness of 82.1% (Actinopterygii). The estimated size of the genome ranges between 655 and 850 Mb. Additionally, we generated a relatively high-level liver transcriptome. To deliver a proof of principle of the value of this dataset, we established the presence and function of enzymes (Elov12, Elov15, and Fads2) that have pivotal roles in the biosynthesis of long chain polyunsaturated fatty acids, essential nutrients particularly abundant in oily fish such as sardines. Our study provides the first

omics dataset from a valuable economic marine teleost species, the European sardine, representing an essential resource for their effective conservation, management, and sustainable exploitation.

Keywords: European sardine; draft genome; teleosts; comparative genomics; long chain polyunsaturated fatty acids

1. Introduction

Teleosts comprise the most species-rich group of vertebrates, with approximately 30,000 described species [1]. During the last decades, teleosts emerged as particularly insightful models for comparative evolutionary studies [2]. Moreover, numerous teleost fish species are of high commercial importance for both fisheries and aquaculture. Fish are not only an important source of protein for human consumers, but oily species represent unique sources of the healthy omega-3 long-chain polyunsaturated fatty acids (LC-PUFAs), which have been shown to have essential roles in cardiovascular health and neuronal development [3,4]. Yet, over-fishing combined with global changes entail countless threats to this taxon, making it an interesting target for aquaculture which already provides approximately half of the seafood consumed worldwide [5]. The oily fish European sardine (*Sardina pilchardus*, Walbaum 1792) (Figure 1) is one of the most commercially important species [6], particularly for the canning industry, and has high nutritional value primarily linked to its omega-3 LC-PUFA content. Interestingly, a steady and sharp decline in capture levels has been observed in the last decade, which is currently dictating severe cuts in fishing quotas within the European Union [7]. Recent advances in captive sardine culture practices represent promising research possibilities [8,9]; however, the complete absence of genomic resources from this iconic species imposes a severe bottleneck. Genome data emerging from high-throughput sequencing technology combined with multiple assembly algorithms has represented a truly transformative event in the field of comparative genomics [10]. Moreover, de novo assemblies based on low coverage and short read approaches are cost effective, and provide valuable biological information [11–13]. Here, we present the first draft genome of the European sardine and provide a relatively high-level liver transcriptome enabling nutrigenomics studies in this iconic species. The genomic makeup of the European sardine was further compared to that of other teleosts, including the closely related clupeid, the Atlantic herring [14]. As a proof of principle, we selected the LC-PUFA biosynthesis, a metabolic pathway accounting for the production of omega-3 fatty acid in fish [15], and characterized key genes with well-established roles within these pathways [3,16].



Figure 1. Photograph of a specimen of European sardine, *Sardina pilchardus* (photograph credits to Mónica Felício and André M. Machado).

2. Methods, Results and Discussion

2.1. Sampling, DNA Extraction, Library Preparation and Genome Sequencing

One *S. pilchardus* specimen (Figure 1) was caught off Esposende (41.501944 N 8.851667 W), Portugal, under the “Programa Nacional de Amostragem Biológica” carried out by the Instituto Português do Mar e da Atmosfera (IPMA). Tissues were harvested immediately and stored in

100% ethanol (muscle) and RNA later (liver) until further processing (Supplementary Table S1 in Supplementary File 2). Genomic DNA was extracted from muscle tissue (~0.5 g) in three replicates, using Qiagen's DNeasy Blood & Tissue Kit (Valencia, CA, USA) according to the manufacturer's instructions, with the following modifications: prior to elution in 100 µL AE buffer, samples were incubated at 37 °C for 10 min, to increase DNA yield. DNA concentration and integrity were verified using an Agilent Genomic DNA ScreenTape (Waldbronn, Germany). We constructed one 150 bp paired-end reads library from 1.2 µg of genomic DNA using the standard Illumina protocol for the TruSeq Nano DNA library kit (Illumina, San Diego, CA, USA). Sequencing was performed with the Illumina HiSeq X Ten (Macrogen, Seoul, Korea) platform and generated 69.0 Gbp of raw reads for downstream analysis (Supplementary Table S1 in Supplementary File 2).

2.2. RNA Extraction, Library Preparation and Sequencing

Total RNA was extracted from liver using Illustra RNAspin Mini RNA Isolation Kit (GE Healthcare, Hammersmith, UK) according to the manufacturer's instructions. The isolated RNA was treated with RNase-free DNase I and eluted with RNase-free water. A strand-specific library with insert size of 250–300 bp was built after conversion of the liver total RNA to cDNA, and sequenced using 150 bp paired-end reads on the Illumina HiSeq 2500 platform by Novogene (Beijing, China). A total of 122.8 million raw reads were produced (Supplementary Table S1 in Supplementary File 2).

2.3. RNA-Seq Raw Data Clean-Up and De Novo Assembly Transcriptome

The quality of raw RNA-Seq reads were scrutinized using the FastQC software (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Trimmomatic (v0.36) [17] was used to clean up the initial dataset (LEADING:15 TRAILING:15 SLIDINGWINDOW:4:20 MINLEN:50) (Table 1). To assemble the paired end reads we used the de novo assembler Trinity (v2.4.0) [18]. All default parameters were used except for SS_lib_type—RF and min_contig_length of 300 bp. To check for contamination sources such as vectors, adapters or other exogenous sequences, we applied the same methodology as previously described [19]. Thus, using the MCSC (Model-based Categorical Sequence Clustering) decontamination pipeline [20] and UniVec database (build 10.0), we obtained the final decontaminated transcriptome assembly further used in the genome annotation process. Overall, from 111.5 million clean reads we obtained a total of 245,053 assembled transcripts, with N50 of 1760 bp (Table 1 and Supplementary Table S1 in Supplementary File 2). Additionally, we also evaluated the gene content completeness using the Benchmarking Universal Single-Copy Orthologs (BUSCO (v.3)) [21]. This analysis was done against the lineage-specific library of *Actinopterygii*, and showed that from 4584 BUSCO ortholog genes, our transcriptome assembly contains 80.6% of the sequences (complete and partial; Table 1).

Table 1. Summary of genome and liver transcriptome statistics of the European sardine, *Sardina pilchardus*.

Features	Genome #	Liver Transcriptome #
Raw Data		
Raw sequencing reads	456,775,568	122,806,922
Clean reads	412,914,751	111,524,231
Contig statistics		
Number of contigs	90,290	245,053
Total contig size, Mb	640.1	278.5
Contig N50 size, bp	10,878	1760
Longest contig, bp	87,474	15,773
GC/AT/N, %	44.45	48.10
Scaffold statistics		
Number of scaffolds	45,321	-
Total scaffold size, Mb	641.5	-
Scaffold N50 size, bp	25,577	-
Longest scaffold, bp	285,113	-
Genome coverage, ×	59	-

Table 1. Cont.

Features	Genome #	Liver Transcriptome #
BUSCO Completeness (Met/Ver/Actino)		
Complete, %	82.7/70.5/68.8	99.1/80.6/72
Complete and single copy, %	78.8/68.4/66.3	41.5/31.2/29.1
Complete and duplicated, %	3.9/2.1/2.5	57.6/49.4/42.9
Fragmented, %	9.2/19.0/13.3	0.6/10.5/8.6
Missing, %	8.1/10.5/17.9	0.3/8.9/19.4
Total BUSCO found	91.9/89.5/82.1	99.7/91.1/80.6
Annotation		
Number of protein-coding genes	29,701	-
Number of functionally annotated proteins	28,783	-
Average CDS length	1561.42	-
Longest CDS	49,643	-
Average protein length	373.45	-
Longest protein	16,525	-
Average number of exon per gene	6.59	-

All statistics are based on contigs/scaffolds of size ≥ 200 bp. Met: From a total of 978 genes of Metazoa library profile; Ver: From a total of 2586 genes of Vertebrata library profile; Actino: From a total of 4584 genes of Actinopterygii library profile.

2.4. DNA Raw Data Clean-Up and Genome Size Estimation

Raw Illumina reads were first processed with Trimmomatic [17] for the removal of adapter sequences, and trimming bases with quality <20 and discarded reads with length <80 . The genome size estimation was performed with two different approaches, using GenomeScope (v1.0.0) [22] and Kmergenie (v1.7044) [23] on genomic clean reads. The first approach requires the Jellyfish (v2.2.6) software to build k-mer frequency distributions. We applied three values of k-mers, 21, 25, and 31, and each histogram was submitted to the GenomeScope software. In the end, we estimated a genome size between 625–637 Mbp, heterozygosity levels between 1.60–1.75%, and unique content of 85.0–85.7% (Supplementary Figure S1A–C in Supplementary File 1 and Supplementary Table S2 in Supplementary File 2). The Kmergenie software with the diploid model was also used, and a genome size of 850 Mb was estimated.

2.5. Assembly and Assessment of Sardine Genome

The genomic clean reads were assembled with the Celera assembler (downloaded from the CVS Concurrent Version System, <http://wgs-assembler.sourceforge.net/>) repository on 21 June 2017; for details see Supplementary Materials and Methods in Supplementary File 1. Interestingly, this assembler has been successful for other teleost species genomes (e.g., *Parachannaichthys charcoti*) [24]. After genome assembly, the clean reads were back-mapped to the sardine genome with BWA mem [25]. PCR duplicates were removed with Picard MarkDuplicates (<http://picard.sourceforge.net>) and local realignment around indels was done with GATK [26]. A median insert size of 441 bp was determined with Picard CollectInsertSizeMetric. To evaluate the genome assembly, we primarily used QUAST v4.3 [27]. Next, the validation of the genome was done with the K-mer analysis toolkit (KAT) [28]. Through this analysis, it was possible to check how the Celera Assembler dealt with the heterozygosity of the sardine. In Supplementary Figure S2 of Supplementary File 1, two peaks can be observed: the first at $25\times$ (heterozygotic) and the second at $55\times$ (homozygotic). Ideally, it is expected that, after the assembly, the shared k-mers contents of both distributions (red zones) are merged (black zones in the first peak), and stayed represented just once in both distributions [28]. Notwithstanding, our distributions show this profile, with the black content of the first peak nearly filling the full area of the second peak. Second, and similarly to the transcriptomic approach, the sardine genome was also inspected in terms of expected gene content with the BUSCO v3 software [21]. From the total of 4584 genes present in the Actinopterygii library, we found 82.1% (complete and fragmented). Importantly, the BUSCO scores are in line with previous reports for other fish assemblies with similar methodologies and sequencing strategies. For instance, Malmström and colleagues [13] reported the genome of 66 teleost species, in which 48 genomes showed between 33 and 84% (complete and

fragmented) of the genes present in the Actinopterygii library. A combination of short and long read provides, however, a higher detection of BUSCO scores, and should be pursued in the near future [29–32]. To further determine the genome completeness, we mapped the de novo assembled liver transcriptome against the genome with Blat [33]. More than 97% of the transcripts have a match hit with at least one genomic scaffold, and 89% of the total number of bases are covered by our assembly (Supplementary Table S3 in Supplementary File 2).

2.6. Genome Annotation

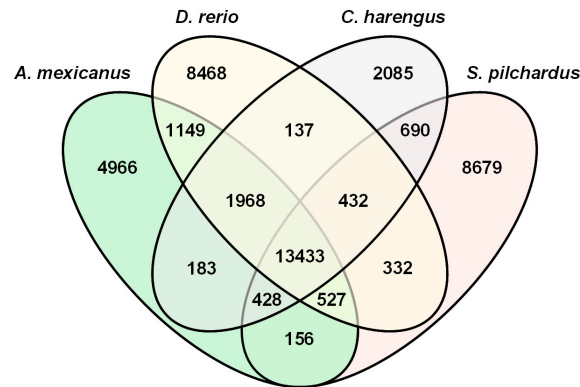
The genome annotation of sardine was performed using two-pass iterative MAKER (v2.31.9) pipeline [34]. Prior to running Maker, we identified repetitive sequences in our genome assembly using an approach described in [35]. Briefly, RepeatModeler (v1.0.8) [36], LTRharvest [37] part of Genome tools (v1.5.7) [38] and TransposonPSI [39] were used in combination to create a set of putative repeats. Elements with a single match against a UniProtKB/SwissProt database and not against the database of known repeated elements included in RepeatMasker were removed. The remaining elements were classified and combined with known repeat elements from RepBase (release 20150807) [40]. Then, the custom repeat database RepBase-derived and RepeatMasker library (release 20150807) [40] were used in the RepeatMasker (v4.0.6) [41] inside the Maker pipeline. In addition to the previously-described Trinity-based transcriptome assembly, the transcriptome reads were mapped to the genome with HISAT (v2.0.5) [42,43] and assembled with StringTie (v1.3.1) [44]. The mapped reads were used to train the GeneMark-ET (v4.32) and AUGUSTUS (v3.2.3) [45] ab initio gene predictors via the tool BRAKER (v1.11) [46]. Splice junctions were detected from the mapped reads with Portcullis (v1.02) [47], and these were used as input to Mikado (v1.2.2) [48], together with the StringTie and Trinity transcriptome assemblies, to merge the redundant transcripts. The resulting GFF file was used as input to Maker (as *est_gff*). The predicted genes from GeneMark-ET were also used as input to Maker (as *pred_gff*), but only in the first iteration, since Maker keeps the gene names as given by GeneMark-ET, and uses them in the output GFF of Maker, which can cause issues with downstream analysis. The first iteration of Maker (which also included proteins from the UniProtKB/SwissProt, cleaned of transposable element proteins) was used to train SNAP (v2013_11_29) [49] with the transcriptome, protein evidence, and custom library of repeats. In the second iteration, we did not use the GeneMark-ET predictions, but utilized AUGUSTUS and SNAP, in addition to the UniProtKB/SwissProt database and the transcriptome.

To functionally annotate the genes and protein models, we used two independent approaches. First, we used a BLAST (v2.6.0) [50] methodology with the following parameters (*blast-p*, *-evalue* 1e−5, *-seg yes*, *-soft masking true*, *-lcase masking*, and *-num_alignments 1*), against the UniProtKB/SwissProt database. In the second approach, we opted for the inclusion of InterProScan [51] searches. The outputs of both approaches were used to refine the gene and protein models, as established in protocols of Campbell et al. (2014) [34]. Finally, our dataset contained 29,701 genes, which were selected based on a maximum Annotation Edit Distance (AED) score of ≤ 0.5 (from 0 to 1, where the 0 corresponds to strong evidence support, and 1 corresponds to no evidence support) (Table 1). The number of predicted coding genes in our dataset is higher than that previously calculated for the *Clupea harengus* genome (23,336 coding gene models) [14]. This discrepancy likely reflects a higher level of gene fragmentation in our assembly, which does not impact the application of the dataset for experimental research (see Section 2.9).

To obtain a broad overview of the annotated gene repertoire in the Clupeidae family, we also compared the ortholog gene collection of sardine with other teleost species including another clupeid, *C. harengus* [14], and to two well-annotated genomes from the sister clade, the Otophysa [2] (for details see Supplementary Materials and Methods in Supplementary File 1). Using the Orthofinder v2.2.6 [52], we identified 24,677 clusters of orthologs genes in the sardine genome: 13,433 orthogroups were shared among the four species, and at least 690 orthogroups were shared exclusively between sardine and herring of the Clupeidae family (Figure 2A). A total of 8679 orthogroups were found to be exclusive to

sardine, with this number likely reflecting gene fragmentation in the assembly process, a consequence of the sequencing strategy.

(A)



(B)

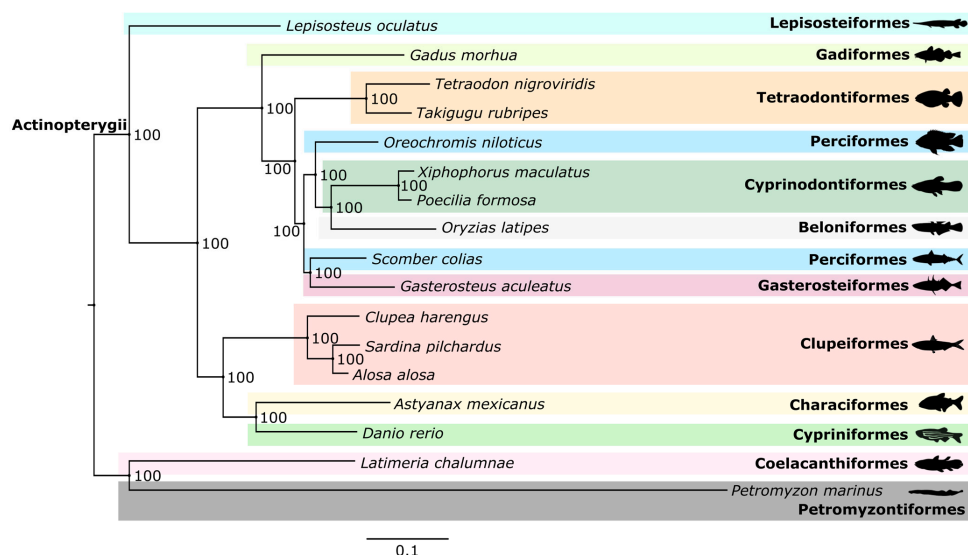


Figure 2. Genome evolution and phylogenomics. (A) Orthologous gene families across four fish genomes (European sardine, zebrafish, herring and blind cave fish). (B) Phylogeny of vertebrates (lamprey as the outgroup species); numbers at nodes represent bootstrap values.

2.7. Sardine Phylogenomics

The phylogenomic analysis was conducted with gene orthologues from 17 fish species, representing 13 orders, providing an ample representation of teleost diversity, with a specific focus in Clupeidae. Transcriptomic sequences for *S. pilchardus*, *C. harengus*, *Alosa alosa*, and *Scomber colias* were clustered based on sequence similarity with zebrafish sequences. Briefly, a blast-p output was filtered to obtain only hits with a percent identity equal or higher than 50% and a length of at least 30 amino acids. We then selected the hit with the highest bitscore. Zebrafish orthologs were retrieved from the ENSEMBL database [53] for all the 17 fish species available, and assigned to the respective cluster. In order to avoid paralogs, only clusters with one sequence per species were considered, resulting in 106 ortholog clusters that included all species (Supplementary File 3). Amino acid sequences were then aligned with MAFFT v7.402 [54] using the model L-INS-i, recommended for a small number of sequences with long gaps. The resulting 106 sequences alignments were then concatenated (42,267 bp long). A maximum-likelihood phylogenetic inference for the concatenated protein alignment was done in ExaML v3 [55], including 100 bootstrap replicates, under the protgammaauto option,

and was computed using parsimony starting trees for each replicate, using RAxML v8.2.12 [56]. In the ExaML tree, two major groups can be observed: one that comprises all Actinopterygian, and another with the Sarcopterygii and Cephalaspidomorphi in the basal position of the tree (Figure 2B). In this tree, all species belonging to the Clupeiformes order are clustered together, and the same for Tetraodontiformes and Cyprinodontiformes. Perciformes are the only order that is separated into two, with *Oreochromis niloticus* closely related to Cyprinodontiformes and *S. colias* with Gasterosteiformes. The position of *Lepisosteus oculatus* at the base of the Actinopterygian cluster was also recovered with maximum support. Overall, our phylogenetic analysis demonstrates the phylogenetic position of the European sardine, together with other clupeids such as the allis shad (*A. alosa*) and the Atlantic herring (Figure 2B). The same general phylogenetic relationships were recovered when the concatenated mitochondrial dataset of protein-coding genes was used. The only exception position was the *A. mexicanus*, in that it was clustered together (with low statistical support) with Clupeiformes and not with zebrafish (Supplementary Materials and Methods and Supplementary Figure S4 of Supplementary File 1, and Supplementary Table S4 in Supplementary File 2).

2.8. Mitochondrial Genome

We used NOVOPlasty (v2.6.5) [57] to perform the de novo assembly of the sardine mitochondrial genome (mtDNA). The assembly was executed using the raw whole genome sequencing reads only with the adapters removed (authors' instructions), and using a *cox1* mtDNA gene nucleotide sequence of the same species (NCBI accession number NC_009592.1 (5484 . . . 7034)). The k-mers length was set to 39, 50, and 75 bp, and all assembly runs resulted in the same mtDNA circular contig with total length of 17,755 bp. We also used NOVOPlasty to detect heteroplasmy in the newly-assembled mtDNA, with a minimum minor allele frequency option set to 0.01 (heteroplasmy detection of >1%). Two heteroplasmic positions were detected in the kmer-75 at positions 3500 (from T to G, alternative allele frequency of 1.23%, depth of coverage of 326, located in *mt-nd1* gene), and another at position 10208 (from T to C, alternative allele frequency of 1.02%, depth of coverage of 391, located in *mt-nad4l* gene). Mitochondrial gene annotations were performed using MITOS (v2) [58] and tRNAs gene limits were rechecked with ARWEN (v1.2) [59]. All typical Metazoan genes were annotated (13 protein coding genes, 22 transfer RNAs, and 2 ribosomal RNAs, Supplementary Figure S3 in Supplementary File 1). The complete mtDNA was deposited in GenBank (Supplementary Table S1 in Supplementary File 2).

2.9. Gene Orthologs of LC-PUFA Desaturation and Elongation Are Present in the Sardine Genome and Transcriptome

To demonstrate the biological value of the omics datasets, we next investigated the key enzymes of LC-PUFA biosynthesis in the sardine draft genome and liver transcriptome, a major metabolic site for PUFA metabolism [15]. More specifically, we investigated the repertoire and function of genes encoding fatty acyl desaturases (*Fads*) and elongation of very long chain fatty acid (*Elovl*) proteins with pivotal roles in LC-PUFA biosynthesis [3]. Among *fads*, our data shows that the European sardine possess one single *fads*-like gene that was confirmed to be orthologous of *fads2* (Figure 3A,C) (for details see Supplementary Materials and Methods in Supplementary File 1 and Supplementary File 4). Our microsynteny analyses shows the conservation of the reconstructed *locus*, when compared to *C. harengus*, and further suggests the absence of *fads1* from sardine's genome, in agreement with the loss of this tandem gene duplicate during teleost evolution (Figure 3C, Supplementary Tables S5 and S6 in Supplementary File 2) [16]. Among *elovl*, we identified two *elovl*-like sequences, namely *elovl2* and *elovl5*, with well-known roles in LC-PUFA biosynthetic pathways (Figure 3B). Again, microsynteny conservation was found in the reconstructed *loci*. While *elovl5* is present in virtually all teleosts [3,60], *elovl2* has only been described in a few of species, and is reported to be lost in the Neoteleostei [60]. Thus, the presence of an *elovl2* gene in sardine is consistent with the phylogenetic location of this species within the Otomorpha group, to which species with characterized *elovl2* belong [61,62]. Accession numbers for the herein isolated *fads2*, *elovl2*, and *elovl5* gene orthologs have been deposited

in GenBank (Supplementary Table S1 in Supplementary File 2), and are located within genome scaffold numbers: *fads2*—scaffolds scf7180014809123 and scf7180014798914; *elovl2*—scaffold scf7180014826570; and *elovl5*—scaffolds scf7180014798588 and scf7180014802726 (for details see Supplementary Materials and Methods in Supplementary File 1 and Supplementary Table S5 in Supplementary File 2).

We next examined the function of the enzymes encoded by the *fads2*, *elovl2*, and *elovl5* genes to establish their contribution to LC-PUFA biosynthesis in sardine using an established yeast-based expression system [63] (Figure 3D) (Supplementary Table S7 in Supplementary File 2; see Supplementary Materials and Methods in Supplementary File 1 for details). The sardine *fads2* encodes a desaturase with $\Delta 6$ and $\Delta 8$ desaturase activities (Supplementary Tables S8 and S9 in Supplementary File 2), typical of vertebrates Fads2 enzymes [3]. Both Elovl2 and Elovl5 were capable of elongating polyunsaturated fatty acids from 18 to 22 carbons, which is consistent with activities reported in other vertebrate orthologs (Supplementary Table S10 in Supplementary File 2). Such enzymatic capabilities enable sardines to produce docosahexaenoic acid (DHA) from eicosapentaenoic acid (EPA, 20:5n-3). However, the lack of $\Delta 5$ desaturation capacity strongly suggests that sardine is unable to produce EPA endogenously or arachidonic acid (ARA, 20:4n-6) (Figure 3D). Importantly, these results clearly illustrate the validity of the herein released omics datasets for nutrigenomic studies.

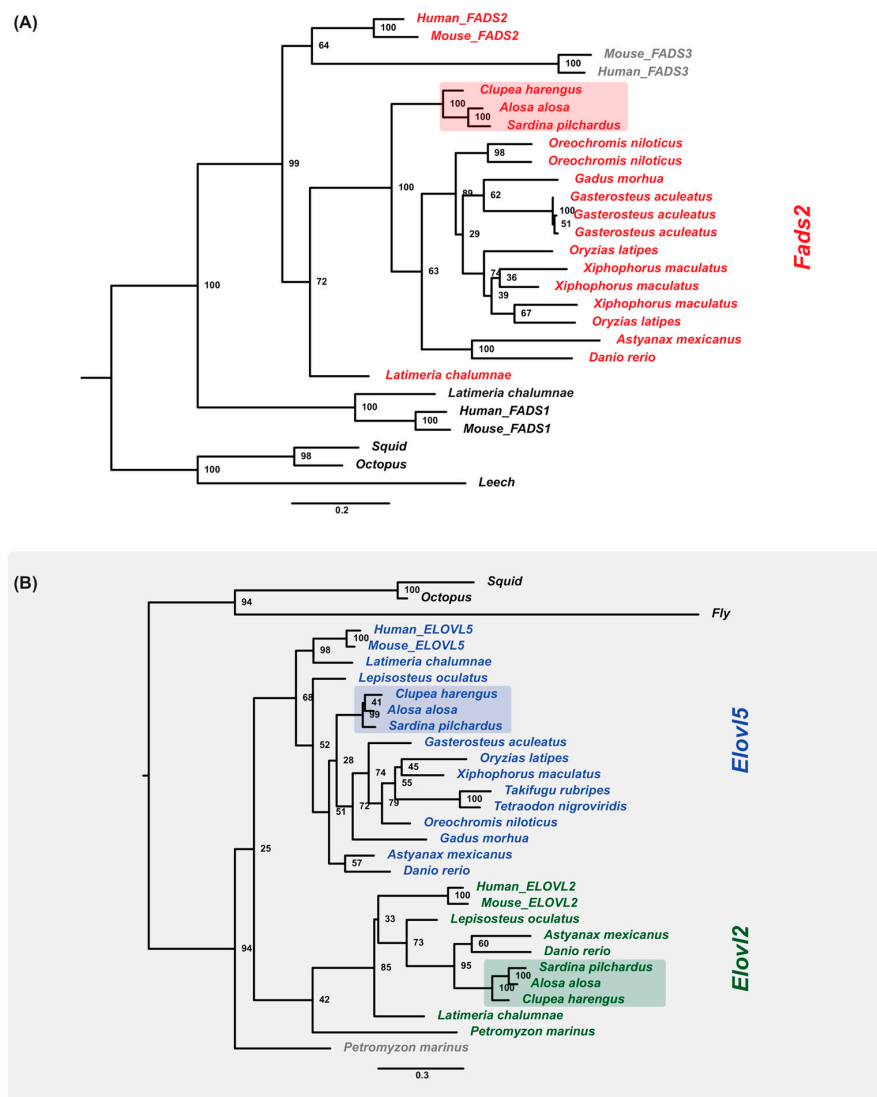


Figure 3. Cont.

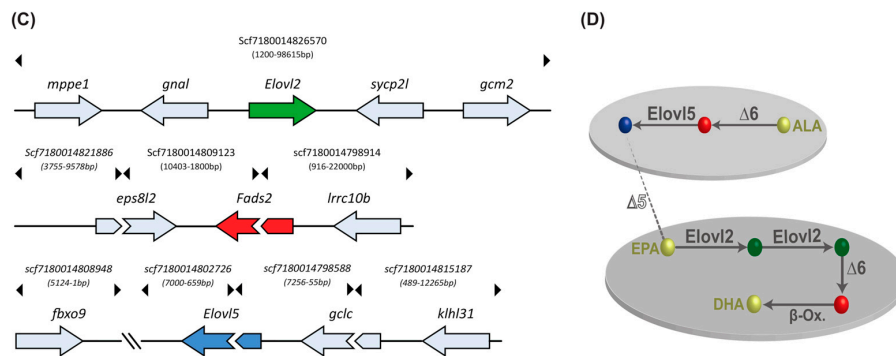


Figure 3. Maximum likelihood phylogenetic analysis of *fads2* (A) and *elovl* orthologues (B) analyzed in the present study: Clupeiformes species are highlighted, node numbers indicate bootstrap values. (C) Reconstructed genomic loci of *fads2*, *elovl2*, and *elovl5* denote syntenic conservation between the European sardine and Atlantic herring: scaffold coordinates and identified neighbouring genes are indicated; broken lines and arrows denote reconstruction from overlapping scaffolds. (D) LC-PUFA biosynthesis pathway in the European sardine, dashed line indicates the $\Delta 5$ desaturation capacity absent in the European sardine, n-3 fatty acids are indicated in yellow: ALA— α -Linolenic acid (18:3n-3), EPA—eicosapentaenoic acid (20:5n-3) and DHA—docosahexaenoic acid (22:6n-3).

3. Conclusions

We generated a draft genome assembly and liver transcriptome of the commercially-important European sardine. We further demonstrate the power of this dataset by exploring the endogenous capacity of sardines (clupeids) to biosynthesize LC-PUFAs. The information retrieved here, and made publicly available, will further contribute not only to elucidating the fundamentals of the physiology, endocrinology, reproduction, and nutrition of sardine, providing an essential framework for future conservation and sustainable exploitation of this iconic species, but will also contribute to future comparative genomic studies, notably regarding life history strategies among teleosts.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4425/9/10/485/s1>, Supplementary File 1 containing Supplementary Figures, Figure S1: Estimation of genome size, repeat content, and heterozygosity by GenomeScope, Figure S2: Assessment of the genome assembly by comparison of read spectrum and assembly copy number, Figure S3: Circular gene map of the mitochondrial genome of *Sardina pilchardus*, Figure S4: Phylogenetic tree of ray-finned fishes estimated from 13 concatenated individual mtDNA protein-coding gene amino acid sequences, Supplementary Materials and Methods, and Supplementary References. Supplementary File 2 containing Supplementary Tables, Table S1: MixS descriptors and accession numbers of tissue samples, raw data and Assemblies of *Sardina pilchardus*, Table S2: GenomeScope profile statistics for *Sardina pilchardus* WGS reads estimated using k-mer 21, 25, and 31, Table S3: Completeness assessment of sardine genome using the liver transcriptome, Table S4: List of species and GenBank references used in the mtDNA phylogeny, Table S5: Blast-n output of *Clupea harengus* sequences against the sardine genome, Table S6: Top results from a reciprocal blast-p search of the 5 proteins flanking the genomic locations of *fads2*, *elovl2* and *elovl5* in herring and the sardine genomes, Table S7: Primer sequences and PCR conditions used to isolate *fads2*, *elovl2* and *elovl5* genes, Table S8: Functional characterization of *Sardina pilchardus* *Elov12* and *Elov15*, Table S9: Functional characterization of *Sardina pilchardus* *Fads2*, Table S10: *Sardina pilchardus* *Fads2* + Zebrafish *Elov12* co-expression. Supplementary File 3 containing Clusters of sequences used to Sardine Phylogenomics Analyses. Supplementary File 4 containing Clusters of sequences used to—Gene orthologs of LC-PUFA desaturation and elongation are present in the sardine genome and transcriptome—analyses. The raw sequencing data (RNA-Seq and WGS), genome assembly, transcriptome assembly, mitochondrial genome and isolated LC-PUFA sequences can be consulted via NCBI. All accession numbers are indicated in Supplementary Table S1 of Supplementary File 2. Supporting data such as protein and transcripts from genome as well as .gff file of the annotation can be obtained from <https://figshare.com/s/98f0644bd974f891143c>.

Author Contributions: Conceptualization, R.R.d.F. and L.F.C.C.; Data curation, A.M.M., O.K.T. and M.F.; Funding acquisition, L.F.C.C.; Investigation, A.M.M., O.K.T., N.K., A.C., B.P., P.F.C., E.F., N.B., M.L.-M., R.F., R.R., M.M.F., S.J., Ó.M., R.R.d.F. and L.F.C.C.; Resources, M.M.F.; Supervision, S.J., Ó.M., R.R.d.F. and L.F.C.C.; Validation, A.M.M. and O.K.T.; Writing—original draft, A.M.M., Ó.M., R.R.d.F. and L.F.C.C.; Writing—review & editing, O.K.T., N.K., A.C., B.P., M.M.F., P.F.C., E.F., N.B., M.L.-M., R.F., R.R., M.M.F. and S.J.

Funding: We acknowledge the North Portugal Regional Operational Program (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF) that supported this research through the Coral—Sustainable Ocean Exploitation (reference NORTE-01-0145-FEDER-000036). R.R.d.F. thanks the Danish National Research Foundation for its support of the Center for Macroecology, Evolution, and Climate (grant DNRF96).

Acknowledgments: Some computational work was performed on the Abel Supercomputing Cluster (Norwegian metacenter for High Performance Computing (NOTUR) and the University of Oslo) operated by the Research Computing Services group at USIT, the University of Oslo IT-department (<http://www.hpc.uio.no/>). We would like to thank Jette Bornholdt, Amal Al-Chaer and George Pacheco for help with laboratory procedures, and the Bioinformatics Center of the University of Copenhagen for providing laboratory space. This work is part of the CIIMAR-lead initiative Portugal-Fishomics.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ravi, V.; Venkatesh, B. The divergent genomes of teleosts. *Annu. Rev. Anim. Biosci.* **2018**, *6*, 47–68. [[CrossRef](#)] [[PubMed](#)]
2. Hughes, L.C.; Ortí, G.; Huang, Y.; Sun, Y.; Baldwin, C.C.; Thompson, A.W.; Arcila, D.; Betancur-R, R.; Li, C.; Becker, L.; et al. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 6249–6254. [[CrossRef](#)] [[PubMed](#)]
3. Castro, L.F.C.; Tocher, D.R.; Monroig, O. Long-chain polyunsaturated fatty acid biosynthesis in chordates: Insights into the evolution of Fads and Elovl gene repertoire. *Prog. Lipid Res.* **2016**, *62*, 25–40. [[CrossRef](#)] [[PubMed](#)]
4. Ghasemifard, S.; Turchini, G.M.; Sinclair, A.J. Omega-3 long chain fatty acid “bioavailability”: A review of evidence and methodological considerations. *Prog. Lipid Res.* **2014**, *56*, 92–108. [[CrossRef](#)] [[PubMed](#)]
5. Food and Agriculture Organization of the United Nations. *The State of World Fisheries And Aquaculture (SOFIA)—Meeting the Sustainable Development Goals*, 1st ed.; FAO: Rome, Italy, 2018.
6. Instituto Nacional de Estatística (INE). *Estatísticas da Pesca—2016*, 1st ed.; Instituto Nacional de Estatística: Lisboa, Portugal, 2017.
7. Silva, A.; Moreno, A.; Riveiro, I.; Santos, B.; Pita, C.; Rodrigues, J.G.; Villasante, S.; Pavlowski, L.; Duhamel, E. *Sardine Fisheries: Resource Assessment and Social and Economic Situation*; European Parliament: Brussels, Belgium, 2015; ISBN 978-92-823-8384-1.
8. Bandarra, N.M.; Marçalo, A.; Cordeiro, A.R.; Pousão-Ferreira, P. Sardine (*Sardina pilchardus*) lipid composition: Does it change after one year in captivity? *Food Chem.* **2018**, *244*, 408–413. [[CrossRef](#)] [[PubMed](#)]
9. Olmedo, M.; Iglesias, J.; Peleteiro, J.; Forés, R.; Miranda, A. Acclimatization and induced spawning of sardine *Sardina pilchardus* Walbaum in captivity. *J. Exp. Mar. Biol. Ecol.* **1990**, *140*, 61–67. [[CrossRef](#)]
10. Fernandez-Silva, I.; Henderson, J.B.; Rocha, L.A.; Simison, W.B. Whole-genome assembly of the coral reef Pearlscale Pygmy Angelfish (*Centropyge vrolikii*). *Sci. Rep.* **2018**, *8*, 1498. [[CrossRef](#)] [[PubMed](#)]
11. Malmstrøm, M.; Matschiner, M.; Tørresen, O.K.; Star, B.; Snipen, L.G.; Hansen, T.F.; Baalsrud, H.T.; Nederbragt, A.J.; Hanel, R.; Salzburger, W.; et al. Evolution of the immune system influences speciation rates in teleost fishes. *Nat. Genet.* **2016**, *48*, 1204–1210. [[CrossRef](#)] [[PubMed](#)]
12. Nakamura, Y.; Mori, K.; Saitoh, K.; Oshima, K.; Mekuchi, M.; Sugaya, T.; Shigenobu, Y.; Ojima, N.; Muta, S.; Fujiwara, A.; et al. Evolutionary changes of multiple visual pigment genes in the complete genome of Pacific bluefin tuna. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 11061–11066. [[CrossRef](#)] [[PubMed](#)]
13. Malmstrøm, M.; Matschiner, M.; Tørresen, O.K.; Jakobsen, K.S.; Jentoft, S. Whole genome sequencing data and de novo draft assemblies for 66 teleost species. *Sci. Data* **2017**, *4*, 160132. [[CrossRef](#)] [[PubMed](#)]
14. Martinez Barrio, A.; Lamichhaney, S.; Fan, G.; Rafati, N.; Pettersson, M.; Zhang, H.; Dainat, J.; Ekman, D.; Höppner, M.; Jern, P.; et al. The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. *Elife* **2016**, *5*, 1–32. [[CrossRef](#)] [[PubMed](#)]
15. Monroig, O.; Tocher, D.R.; Castro, L.F.C. Polyunsaturated Fatty Acid Biosynthesis and Metabolism in Fish. In *Polyunsaturated Fatty Acid Metabolism*, 1st ed.; Burdige, G., Ed.; AOCS Press: Urbana, IL, USA, 2018.

16. Castro, L.F.C.; Monroig, Ó.; Leaver, M.J.; Wilson, J.; Cunha, I.; Tocher, D.R. Functional desaturase fads1 ($\Delta 5$) and fads2 ($\Delta 6$) orthologues evolved before the origin of jawed vertebrates. *PLoS ONE* **2012**, *7*, e31950. [[CrossRef](#)] [[PubMed](#)]
17. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
18. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [[CrossRef](#)] [[PubMed](#)]
19. Machado, A.M.; Felício, M.; Fonseca, E.; da Fonseca, R.R.; Castro, L.F.C. A resource for sustainable management: De novo assembly and annotation of the liver transcriptome of the Atlantic chub mackerel, *Scomber colias*. *Data Brief* **2018**, *18*, 276–284. [[CrossRef](#)] [[PubMed](#)]
20. Lafond-Lapalme, J.; Duceppe, M.O.; Wang, S.; Moffett, P.; Mimee, B. A new method for decontamination of de novo transcriptomes using a hierarchical clustering algorithm. *Bioinformatics* **2017**, *33*, 1293–1300. [[CrossRef](#)] [[PubMed](#)]
21. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)] [[PubMed](#)]
22. Vurture, G.W.; Sedlazeck, F.J.; Nattestad, M.; Underwood, C.J.; Fang, H.; Gurtowski, J.; Schatz, M.C. GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics* **2017**, *33*, 2202–2204. [[CrossRef](#)] [[PubMed](#)]
23. Chikhi, R.; Medvedev, P. Informed and automated k-mer size selection for genome assembly. *Bioinformatics* **2014**, *30*, 31–37. [[CrossRef](#)] [[PubMed](#)]
24. Ahn, D.H.; Shin, S.C.; Kim, B.M.; Kang, S.; Kim, J.H.; Ahn, I.; Park, J.; Park, H. Draft genome of the Antarctic dragonfish, *Parachaenichthys charcoti*. *Gigascience* **2017**, *6*, 1–6. [[CrossRef](#)] [[PubMed](#)]
25. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)] [[PubMed](#)]
26. Depristo, M.A.; Banks, E.; Poplin, R.; Garimella, K.V.; Maguire, J.R.; Hartl, C.; Philippakis, A.A.; Del Angel, G.; Rivas, M.A.; Hanna, M.; et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **2011**, *43*, 491–501. [[CrossRef](#)] [[PubMed](#)]
27. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [[CrossRef](#)] [[PubMed](#)]
28. Mapleson, D.; Accinelli, G.G.; Kettleborough, G.; Wright, J.; Clavijo, B.J. KAT: A K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics* **2017**, *33*, 574–576. [[CrossRef](#)] [[PubMed](#)]
29. Austin, C.M.; Tan, M.H.; Harrisson, K.A.; Lee, Y.P.; Croft, L.J.; Sunnucks, P.; Pavlova, A.; Gan, H.M. De novo genome assembly and annotation of Australia’s largest freshwater fish, the Murray cod (*Maccullochella peelii*), from Illumina and Nanopore sequencing read. *Gigascience* **2017**, *6*. [[CrossRef](#)] [[PubMed](#)]
30. Kim, H.S.; Lee, B.Y.; Han, J.; Jeong, C.B.; Hwang, D.S.; Lee, M.C.; Kang, H.M.; Kim, D.H.; Lee, D.; Kim, J.; et al. The genome of the marine medaka *Oryzias melastigma*. *Mol. Ecol. Resour.* **2018**, *18*, 656–665. [[CrossRef](#)] [[PubMed](#)]
31. Tan, M.H.; Austin, C.M.; Hammer, M.P.; Lee, Y.P.; Croft, L.J.; Gan, H.M. Finding Nemo: Hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (*Amphiprion ocellaris*) genome assembly. *Gigascience* **2018**, *7*, 1–6. [[CrossRef](#)] [[PubMed](#)]
32. Marcionetti, A.; Rossier, V.; Bertrand, J.A.M.; Litsios, G.; Salamin, N. First draft genome of an iconic clownfish species (*Amphiprion frenatus*). *Mol. Ecol. Resour.* **2018**, *18*, 1092–1101. [[CrossRef](#)] [[PubMed](#)]
33. Kent, W.J. BLAT—The BLAST-like alignment tool. *Genome Res.* **2002**, *12*, 656–664. [[CrossRef](#)] [[PubMed](#)]
34. Campbell, M.S.; Holt, C.; Moore, B.; Yandell, M. Genome Annotation and Curation Using MAKER and MAKER-P. *Curr. Protoc. Bioinform.* **2014**, *2014*, 4.11.1–4.11.39. [[CrossRef](#)]
35. Tørresen, O.K.; Star, B.; Jentoft, S.; Reinart, W.B.; Grove, H.; Miller, J.R.; Walenz, B.P.; Knight, J.; Ekholm, J.M.; Peluso, P.; et al. An improved genome assembly uncovers prolific tandem repeats in Atlantic cod. *BMC Genom.* **2017**, *18*, 95. [[CrossRef](#)] [[PubMed](#)]
36. Price, A.L.; Jones, N.C.; Pevzner, P.A. De novo identification of repeat families in large genomes. *Bioinformatics* **2005**, *21*, i351–i358. [[CrossRef](#)] [[PubMed](#)]

37. Ellinghaus, D.; Kurtz, S.; Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinform.* **2008**, *9*, 18. [[CrossRef](#)] [[PubMed](#)]
38. Gremme, G.; Steinbiss, S.; Kurtz, S. Genome tools: A comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2013**, *10*, 645–656. [[CrossRef](#)] [[PubMed](#)]
39. TransposonPSI: An Application of PSI-Blast to Mine (Retro-)Transposon ORF Homologies. Available online: <http://transposonpsi.sourceforge.net/> (accessed on 18 April 2018).
40. Jurka, J.; Kapitonov, V.V.; Pavlicek, A.; Klonowski, P.; Kohany, O.; Walichiewicz, J. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **2005**, *110*, 462–467. [[CrossRef](#)] [[PubMed](#)]
41. Tarailo-Graovac, M.; Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* **2009**, *25*, 4.10.1–4.10.14. [[CrossRef](#)]
42. Pertea, M.; Kim, D.; Pertea, G.M.; Leek, J.T.; Salzberg, S.L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **2016**, *11*, 1650–1667. [[CrossRef](#)] [[PubMed](#)]
43. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [[CrossRef](#)] [[PubMed](#)]
44. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.C.; Mendell, J.T.; Salzberg, S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **2015**, *33*, 290–295. [[CrossRef](#)] [[PubMed](#)]
45. Stanke, M.; Steinkamp, R.; Waack, S.; Morgenstern, B. AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Res.* **2004**, *32*, W309–W312. [[CrossRef](#)] [[PubMed](#)]
46. Hoff, K.J.; Lange, S.; Lomsadze, A.; Borodovsky, M.; Stanke, M. BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **2015**, *32*, 767–769. [[CrossRef](#)] [[PubMed](#)]
47. Mapleson, D.; Venturini, L.; Kaithakottil, G.; Swarbreck, D. Efficient and accurate detection of splice junctions from RNAseq with Portcullis. *bioRxiv* **2017**, 217620. [[CrossRef](#)]
48. Venturini, L.; Caim, S.; Kaithakottil, G.G.; Mapleson, D.L.; Swarbreck, D. Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *Gigascience* **2018**, *7*, giy093. [[CrossRef](#)] [[PubMed](#)]
49. Grenon, P.; Smith, B. SNAP and SPAN: Towards dynamic spatial ontology. *Spat. Cognit. Comput.* **2004**, *4*, 69–104. [[CrossRef](#)]
50. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
51. Quevillon, E.; Silventoinen, V.; Pillai, S.; Harte, N.; Mulder, N.; Apweiler, R.; Lopez, R. InterProScan: Protein domains identifier. *Nucleic Acids Res.* **2005**, *33*, W116–W120. [[CrossRef](#)] [[PubMed](#)]
52. Wang, Y.; Coleman-Derr, D.; Chen, G.; Gu, Y.Q. OrthoVenn: A web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* **2015**, *43*, W78–W84. [[CrossRef](#)] [[PubMed](#)]
53. Kersey, P.J.; Allen, J.E.; Allot, A.; Barba, M.; Boddu, S.; Bolt, B.J.; Carvalho-Silva, D.; Christensen, M.; Davis, P.; Grabmueller, C.; et al. Ensembl Genomes 2018: An integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Res.* **2018**, *46*, D802–D808. [[CrossRef](#)] [[PubMed](#)]
54. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [[CrossRef](#)] [[PubMed](#)]
55. Kozlov, A.M.; Aberer, A.J.; Stamatakis, A. ExaML version 3: A tool for phylogenomic analyses on supercomputers. *Bioinformatics* **2015**, *31*, 2577–2579. [[CrossRef](#)] [[PubMed](#)]
56. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)] [[PubMed](#)]
57. Dierckxsens, N.; Mardulyn, P.; Smits, G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **2016**, *45*, e18. [[CrossRef](#)]
58. Bernt, M.; Donath, A.; Jühling, F.; Externbrink, F.; Florentz, C.; Fritzsch, G.; Pütz, J.; Middendorf, M.; Stadler, P.F. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* **2013**, *69*, 313–319. [[CrossRef](#)] [[PubMed](#)]
59. Laslett, D.; Canbäck, B. ARWEN: A program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* **2008**, *24*, 172–175. [[CrossRef](#)] [[PubMed](#)]

60. Monroig, Ó.; Lopes-Marques, M.; Navarro, J.C.; Hontoria, F.; Ruivo, R.; Santos, M.M.; Venkatesh, B.; Tocher, D.R.C.; Castro, L.F. Evolutionary functional elaboration of the Elovl2/5 gene family in chordates. *Sci. Rep.* **2016**, *6*, 20510. [[CrossRef](#)] [[PubMed](#)]
61. Morais, S.; Monroig, O.; Zheng, X.; Leaver, M.J.; Tocher, D.R. Highly unsaturated fatty acid synthesis in Atlantic salmon: Characterization of ELOVL5- and ELOVL2-like elongases. *Mar. Biotechnol.* **2009**, *11*, 627–639. [[CrossRef](#)] [[PubMed](#)]
62. Oboh, A.; Betancor, M.B.; Tocher, D.R.; Monroig, O. Biosynthesis of long-chain polyunsaturated fatty acids in the African catfish *Clarias gariepinus*: Molecular cloning and functional characterisation of fatty acyl desaturase (*fads2*) and elongase (*elovl2*) cDNAs. *Aquaculture* **2016**, *462*, 70–79. [[CrossRef](#)]
63. Kabeya, N.; Yevzelman, S.; Oboh, A.; Tocher, D.R.; Monroig, O. Essential fatty acid metabolism and requirements of the cleaner fish, ballan wrasse *Labrus bergylta*: Defining pathways of long-chain polyunsaturated fatty acid biosynthesis. *Aquaculture* **2018**, *488*, 199–206. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).